# CP2403 - Project – Part 2 - Regression

First Name: Kaung Khant

Last Name: Naing

|  |
| --- |
| **1: Scatter plots between each explanatory variable and response variable** |
| 1. **Scatterplot for the association between pressure decibars and reported depth**   There is a positive and strong relationship between Pressure in decibars and Reported depth. Moreover, they have linear correlation.     1. **Scatterplots for the association between reported temperature and reported depth**   Based on the data it can be seen that Reported temperature and Reported depth has a strong and negative relationship, moreover they have a nonlinear correlation between the two data |
| **2: List all the explanatory variables selected for regression analysis. Justify your selection** |
| association between Pressure in decibars and Reported Depth  (0.9999996497614666, 0.0)  association between Reported Temperature of Water and Reported Depth  (-0.9409113898216722, 5.218724267059692e-71)  The chosen data from CalCOFI database for this analysis are Pressure in decibars, reported temperature and reported depth. The reason we selected these variables is because we are investigating the reason of the collapse of sardine population in California sea territory by seeking to find the relationship between the temperature as well as pressure of the sea water with the reported depth in that territory in order to know if they are affecting the sardine population in the area. |
| **3: Regression analysis results** |
| 1. **Regression analysis of Reported Depth and Pressure in decibars**      1. **Regression analysis of Reported Depth and Reported Temperature** |
| **4: Regression equation/line** |
| 1. **Regression equation/line for pressure in decibars and reported depth**   Regression equation for the association between Pressure in decibars and reported depth  b = 0.3712  m = 0.9910  Reported depth = 0.3712 + 0.9910 (pressure in decibars).       1. **Regression equation/line for reported temperature and reported depth**   Regression equation for association between reported temperature and reported depth:  b = 1272.8087  m = -128.7966  R­eported depth = 1272.8087 + (-128.7966) (reported temperature). |
| **5: qqplot** |
|  |
| **6: Conclusion from qqplot** |
| Based on the data shown above it can be concluded that there are some outliers that can  affected the data as a whole which indicates the data are not normally distributed |
| **7: percentage of observations over 2 standardized deviation** |
| 3.3557046979865772% |
| **8: percentage of observations over 2.5 standardized** |
| 2.013422818791946% |
| **9: Conclusion from observations over 2 std and 2.5 std** |
| Based on the information above it can be concluded that because 3.36% of observation  have 2.0 standardize residual and 2% have 2.5 standardize residual, it means that the data is poor fit because there are observations over 2 which indicates that there are outliers. |